







OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002

TIME: 11:36:15

Input Set : N:\Crf3\RULE60\10007521.raw Output Set: N:\CRF3\02042002\J007521.raw

SEQUENCE LISTING

```
(1) GENERAL INFORMATION:
             (i) APPLICANT: Schulein, Martin
     5
                            Andersen, Lene N.
     6
                            Lassen, Soren F.
     7
                            Kauppinen, Markus S.
                                                          ENTERED
     8
                            Lange, Lene
     9
                            Nielsen, Ruby I.
     10
                            Ihara, Michiko
     11
                            Takagi, Shinobu
     12
            (ii) TITLE OF INVENTION: Novel Endoglucanases
     14
           (iii) NUMBER OF SEQUENCES: 109
     16
            (iv) CORRESPONDENCE ADDRESS:
     18
                  (A) ADDRESSEE: Novo Nordisk of North America, Inc.
     19
                  (B) STREET: 405 Lexington Avenue, 64th Floor
     20
                  (C) CITY: New York
     21
                  (D) STATE: New York
     22
                  (E) COUNTRY: United States of America
     23
                  (F) ZIP: 10174-6401
     24
             (V) COMPUTER READABLE FORM:
     26
                  (A) MEDIUM TYPE: Floppy disk
     27
                  (B) COMPUTER: IBM PC compatible
     28
                  (C) OPERATING SYSTEM: PC-DOS/MS-DOS
     29
                  (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
     30
            (vi) CURRENT APPLICATION DATA:
     32
                   (A) APPLICATION NUMBER: US/10/007,521
C--> 33
                   (B) FILING DATE: 10-Dec-2001
C-->34
     35
                  (C) CLASSIFICATION:
           (vii) PRIOR APPLICATION DATA:
     38
                   (A) APPLICATION NUMBER: US/08/651,136
     39
                   (B) FILING DATE: 21-MAY-1996
     40
          (viii) ATTORNEY/AGENT INFORMATION:
     43
                   (A) NAME: Lambiris, Elias J.
     44
                   (B) REGISTRATION NUMBER: 33,728
     45
                   (C) REFERENCE/DOCKET NUMBER: 4366.200-US
     46
            (ix) TELECOMMUNICATION INFORMATION:
     48
                   (A) TELEPHONE: 212-867-0123
     49
                   (B) TELEFAX: 212-878-9655
     50
     53 (2) INFORMATION FOR SEQ ID NO: 1:
             (i) SEQUENCE CHARACTERISTICS:
     55
                   (A) LENGTH: 960 base pairs
     56
                   (B) TYPE: nucleic acid
     57
                   (C) STRANDEDNESS: single
     58
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002 TIME: 11:36:15

59 61 63 64		•) MOI) FEA	LECUI ATURI		PE:	CDNA										
65			•	•	CATI			.787	,								
67																	
																	60
	GCCT	TTG	ATA A	ACCCC	CTCC	cc cc	SACTO	CAGTO	AAC	SACGA	ACGC	ATAC	CTTG	GCA C	CC AT	ľG	115
72									•						Me		
73																1	
	CAT																163
	His	Leu	Ser	_	Thr	Thr	Gly	Phe		Ala	Leu	Pro	Val		Ala	Leu	
77	03.0	a. a	ama	5	000		000	a. a	10		000	m > 0	maa	15	maa	maa	011
	GAC					-											211
81	Asp	GIII	20	ser	СТА	rre	GLY	25	Thr	THE	Arg	туг	_	ASP	Cys	Cys	
	አአሮ	ccc		TICC	CCC	TICC.	ccc		770	ccc	ccc	mac	30	CCC	CMC	CAC	259
	AAG Lys																259
85	цуъ	35	ser	Cys	ніа	пр	40	GTĀ	пуъ	GTÅ	PIO	. 45	ser	PIO	vai	GIII	
	GCC		CAC	AAG	ልልሮ	CAC	-	CCG	СТС	AAC	CAC		ccc	TCC	አርር	CGG	307
	Ala																
89	50	C 15	тор	בינה	11511	55	11511	110	пси	ASH	60	OLY.	GIY	ber	1111	65	
	TCC	GGC	TGC	GAC	GCG		GGC	AGC	GCC	TAC		TGC	TCC	TCC	CAG		355
	Ser																333
93		1	-1-	E	70	V-1	0 -1			75		0,10		501	80	001	
95	CCC	TGG	GCC	GTC		GAC	GAG	CTG	TCG		GGC	TGG	GCG	GCC		AAG	403
	Pro																
97		_		85		-			90	-	•	•		95		-	
99	CTC	GCC	GGC	AGC	TCC	GAG	TCG	CAG	TGG	TGC	TGC	GCC	TGC	TAC	GAG	CTG	451
100	Leu	Ala	Gly	Ser	Ser	Glu	Ser	Gln	Trp	Cys	Cys	Ala	Cys	Tyr	Glu	Leu	
101			100		•			105					110)			
103	ACC	TTC	ACC	AGC	GGG	CCG	GTC	GCG	GGC	AAG	AAG	ATG	ATI	GTG	CAG	GCG	499
104	Thr	Phe	Thr	Ser	Gly	Pro	Val	Ala	Gly	Lys	Lys	Met	Ile	val	Gln	Ala	
105		115					120					125					
																ATC	547
		Asn	Thr	Gly	Gly			Gly	Asp	Asn			Asp	Leu	Ala	Ile	
	130					135					140					145	
																GGC	595
		GIY	, GTA	GIA		_	Ile	Phe	Asn		_	Thr	Asp	GIn	_	Gly	
113				330	150		000	03.0	000	155			3.000		160		642
																AÁG	643
117		PIC	PLO	165		тгр	GIY	ASP	170	-	GIY	GIY	тте			Lys	
		GAG	TCC			ጥጥር	CCC	CAC			אאר	000	ccc	175		TGG	691
																Trp	091
121		GIU	180		SET	FIIE	FIU	185		ьeu	пуб	FIO	190		ווכח	TIP	
	CGC	ጥጥር			ጥጥር	CAA	AAC			AAC	ככב	<u> </u>			ጥጥር	CAG	739
																Gln	, 55
125		195		F	_ 55.5		200					205					
							. •										

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002 TIME: 11:36:15

127 GAG GTG GCC TGC CGT GCG GAG CTC ACG TCC AAG AGC GGC TGC TCC CGT 128 Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser Arg 215 220 225 22														000	mca	mcc	ССТ		787
129 210 215 220 223 243	127	GAG	GTG	GCC	TGC	CCG	TCG	GAG	CTC	ACG	TCC	AAG	AGC	Clv	Cuc	Ser	CG1 Δrσ		, , ,
129 210 213 TARGAGGGAA GAGAGGGGG TGGAAGGACC GAAAGATTCA ACCTCTGCT CTGCTGGGGA 847 133 AGCTCGGCGG CGAGTGTGAA ACTGGTGTAA ATATTGTGGC ACACACAAGC TACTACAGTC 907 966 138 (2) INFORMATION FOR SEO ID NO: 2: 140			Val	Ala	Cys			Glu	Leu	Thr	ser	TÀS	ser	GIÀ	Cys	361	225		
131 TARGAGGAA GACAGGGGC TOGAA ACTGGTGTAA ATATTGTGC ACACACAGC TACTACAGTC 132 AGCTCCGGCC GCGGGTTACA ACTGGTGTAA ATATTGTGC ACACACAGC TACTACAGTC 133 (2) INFORMATION FOR SED ID NO: 2: 140	129	210					215		C 2 C C	C 3 3			አሮሮሞ	ירייני	ጥሮ ር				847
135 CONTINUEDRAL CONTINUEDRAL	131	TAAG	AGGG	AA G	AGAG	GGGG	C TG	GAAG	GACC	GAA	₩₩₩₽ ₩₩₩₩	CCC	ACCI	ממחמי	GC T	ACTA	CAGTC		
138 (2) INFORMATION FOR SEQ ID NO: 2: 140	133	CCTCTCCCCC TCCGCCTAAC TAGCCTTGCT GCGGATCTGT CCAAAAAAAA AAA 96															960		
(i) SEQUENCE CHARACTERISTICS: 141	135	CGTC	TCGC	CG T	CCGG	CTAA	C TA	TD N	1001		GAIC	.101	CCAA						
141 (A) LENGTH: 225 amino acids 142 (B) TYPE: amino acids 143 (D) TOPOLOGY: linear 145 (ii) MOLECULE TYPE: protein 147 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2: 149 Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala 150 1 5 10 15 152 Leu Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys 153 20 25 30 155 Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val 156 35 40 45 158 Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr 159 50 55 60 161 Arg Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln 162 65 70 75 80 164 Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val 165 85 90 95 167 Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu 168 100 105 110 170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln 171 115 120 120 173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala 174 130 135 176 Tle Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145 150 150 155 160 180 180 180 180 185 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 180 180 180 180 185 180 180 180 180 180 180 180 180 180 180			INFC	RMAT	TON	אטע ידי כיני	אסאר סבע	א טב זמשחי	יט. ב כידר	·									
142 (B) TYPE: amino acid 143 (D) TOPOLOGY: linear 145 (ii) MOLECULE TYPE: protein 147 (xi) SEQUENCE DESCRIPPION: SEQ ID NO: 2: 149 Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala 150 1 5 10 15 152 Leu Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys 153 20 25 30 155 Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val 156 35 40 45 158 Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr 159 50 161 Arg Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln 162 65 70 70 75 164 Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val 165 85 167 Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu 168 100 105 170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln 171 115 173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala 174 130 135 140 175 167 148 Pro Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145 150 150 180 165 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 180 185 180 180 180 185 180 180 185 180 180 180 185 180 180 185 180 180 180 185 180 180 180 185 180 180 180 185 180 180 180 185 180 180 185 180 180 180 185 180 180 180 185 180 180 180 185 180 180 180 185 180 180 180 180 180 181 Arg 182 225 185 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 188 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single			(A) LENGTH: 225 amino acids																
143										ucro									
145																			
147			/ii\							٠									
149 Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Val Leu Ala 150	1/7		/xil	SEC	MENC	E DE	ESCRI	PTIC	N: S	EQ I	D NC): 2:	:						
150	147	Met	His	Leu	Ser	Ala	Thr	Thr	Gly	Phe	Leu	Ala	Leu	Pro	Val	Leu	Ala		
152 Leu Asp Gln Leu Ser Gly Ile Gly Gln Thr Thr Arg Tyr Trp Asp Cys 153	150	1				5					10					13			
153	152	Leu	Asp	Gln	Leu	Ser	Gly	Ile	Gly	Gln	Thr	Thr	Arg	Tyr	Trp	Asp	Cys		
155 Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val 156	153				20					25					30				
156	155	Cys	Lys	Pro	Ser	Cys	Ala	Trp	Pro	Gly	Lys	Gly	Pro	Ser	Ser	Pro	Val		
159	156			35					40					45					
159	158	Gln	Ala	Cys	Asp	Lys	Asn	Asp	Asn	Pro	Leu	Asn	Asp	Gly	Gly	Ser	Thr		
162 65 164 Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val 165 85 90 90 95 167 Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu 168 100 105 110 170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln 171 115 120 125 173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala 174 130 135 140 176 Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145 150 155 160 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 165 170 175 182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 183 180 180 185 185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186 195 200 205 188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 115 SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	150		50					55					60						
162 65 164 Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val 165 85 90 90 95 167 Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu 168 100 105 110 170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln 171 115 120 125 173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala 174 130 135 140 176 Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145 150 155 160 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 165 170 175 182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 183 180 180 185 185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186 195 200 205 188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 115 SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	161	Arg	Ser	Gly	Cys	Asp	Ala	Gly	Gly	Ser	Ala	Tyr	Met	Cys	Ser	ser	GIN		
165	162	65					70					75					80		
165	164	Ser	Pro	Trp	Ala		Ser	Asp	Glu	Leu	Ser	Tyr	GIA	Trp	Ala	Ala	Val		
168 170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln 171 115 120 125 173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala 174 130 135 140 176 Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145 150 150 155 160 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 165 170 170 175 182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 180 180 185 185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186 195 200 205 188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 215 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) Type: nucleic acid 200 (C) STRANDEDNESS: single	165					85		_	_			a	2	310	Crrc		Clu		
170 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln 171 115 120 120 125 173 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala 174 130 135 140 176 Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145 150 150 160 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 165 170 170 175 182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 183 180 185 185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186 195 200 205 188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 215 220 191 Arg 192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) Type: nucleic acid 200 (C) STRANDEDNESS: single	167	Lys	Leu	Ala		Ser	Ser	Glu	Ser	GIn	Trp	Cys	Cys	Ald	110	тĂт	Giu		
171	168				100			_			a1	T	T	Mot		Va l	Gln		
171	170	Leu	Thr		Thr	Ser	GŢĀ	Pro	val	ALA	GTA	гÀг	гур	125	TIE	VUI	GIII		
174	171	_		115	1 ·	a1	01	3 an		C1**	λcn	λen	Hic		Asp	Leu	Ala		
176 Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr 177 145				Asn	Thr	GTĀ	GIY		Leu	GLY	АЗР	ASII	140	1110	P	200			
177 145 179 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser 180 165 170 182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 183 180 185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186 195 200 188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 215 220 191 Arg 192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	174	71.	T30	C1	C1.	C1.,	17 a 1	C122	τlο	Phe	Asn	Ala		Thr	Asp	Gln	Tyr		
177				СТА	СТА	GIY	150	GLY	110	1110		155	-1-		-		160		
180 182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 183 180 185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186 195 200 205 188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 215 220 191 Arg 192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	170	C1.	712	Dro	Dro	Δen	Glv	Trp	Glv	Asp	Arg			Gly	Ile	His	Ser		
182 Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn 183			нта	FIU	110	165	O ₁	115	0-1		170	•	•	-		175			
183	197	T.ve	Glu	Glu	Cvs	Glu	Ser	Phe	Pro	Glu	Ala	Leu	Lys	Pro	Gly	Cys	Asn		
185 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe 186	183				180					185					190				
186	185	Tro	Ara	Phe	Asp	Trp	Phe	Gln	Asn	Ala	Asp	Asn	Pro	Ser	Val	Thr	Phe		
188 Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser 189 210 215 220 191 Arg 192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	186			195					200					205					
189 210 215 220 191 Arg 192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	188	Gln	Glu	Val	Ala	Cys	Pro	Ser	Glu	Leu	Thr	Ser	Lys	Ser	Gly	Cys	Ser		
191 Arg 192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single						-		215					220	•					
192 225 195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single																			
195 (2) INFORMATION FOR SEQ ID NO: 3: 197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	192	225																	
197 (i) SEQUENCE CHARACTERISTICS: 198 (A) LENGTH: 894 base pairs 199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single	195	(2)	INF	ORMA	TION	FOR	SEQ	ID	NO:	3 :									
199 (B) TYPE: nucleic acid 200 (C) STRANDEDNESS: single			(i) SE	QUEN	CE C	HARA	CTER	ISTI	CS:									
200 (C) STRANDEDNESS: single				(A) L	ENGT	н: 8	94 b	ase	pair	S								
				(в) т	YPE:	nuc	leic	aci	ď			•						
201 (D) TOPOLOGY: linear				(C) S	TRAN	DEDN	ESS:	sin	ате									
	201			(D) T	OPOL	OGY:	lln	ear										

RAW SEQUENCE LISTING

DATE: 02/04/2002 TIME: 11:36:15

PATENT APPLICATION: US/10/007,521

	203		(ii)	моі	LECUI	LE TY	YPE:	cDN2	A										
	205		(ii) MOLECULE TYPE: (ix) FEATURE:																
	206			(!	A) NA	AME/E	KEY:	CDS											
	207			į.	3) L(CAT:	ON:	1891											
	209		(xi)) SE	QUENC	CE DI	ESCR	PTIC	PTION: SEQ ID NO: 3:										
	211	ATG	CAT	CTC	TCC	GCC	ACC	ACC	GGG	TTC	CTC	GCC	CTC	CCG	GTC	CTG	GCC	48	
									Gly										
W>	213					230					235					240			
	215	CTG	GAC	CAG	CTC	TCG	GGC	ATC	GGC	CAG	ACG	ACC	CGG	TAC	TGG	GAC	TGC	96	
	216	Leu	Asp	Gln	Leu	Ser	Gly	Ile	Gly	Gln	Thr	Thr	Arg	Tyr	Trp	Asp	Cys		
W>	217				245					250					255				
									CCC									144	
	220	Cys	Lys	Pro	Ser	Cys	Ala	Trp	Pro	Gly	Lys	Gly	Pro	Ser	Ser	Pro	Val		
W>				260					265					270					
									AAC									192	
	224	Gln	Ala	Cys	Asp	Lys	Asn	Asp	Asn	Pro	Leu	Asn	Asp	Gly	Gly	Ser	Thr		
W>			275					280					285						
									GGC									240	
	228	Arg	Ser	Gly	Cys	Asp	Ala	Gly	Gly	Ser	Ala	Tyr	Met	Cys	Ser	Ser			
M>							295					300					305		
									GAG									288	
	232	Ser	Pro	Trp	Ala	Val	Ser	Asp	Glu	Leu	Ser	Tyr	Gly	\mathtt{Trp}	Ala		Val		
M>						310					315					320			
									TCG									336	
	236	Lys	Leu	Ala	Gly	Ser	Ser	Glu	Ser		Trp	Cys	Cys	Ala		Tyr	Glu	•	
M>					325					330					335			204	
									GTC									384	
	240	Leu	Thr		Thr	Ser	Gly	Pro	Val	Ala	Gly	Lys	Lys		He	vaı	GIn		
M>				340					345					350	a a	от о		422	
									CTG									432	
		Ala		Asn	Thr	Gly	GLY		Leu	GTÄ	Asp	Asn		Pne	Asp	ьeu	Ala		
W>			355					360	3 mm	mm.a		222	365	3.00	C 3 C	CAC	mac	480	
									ATT									400	
			Pro	GIY	GLY	GIY		GIY	Ile	Pne	ASII		.Cys	TIII	ASP	GIII	385		
M>				000	000		375	mcc	cca	CAC	aca	380	ccc	ccc	አሞር	СУП		528	
									GGC									320	
7.7		СТА	Ата	PIO	PIO	390	СТА	тъ	Gly	изр	395	ıyı	СТУ	GIY	116	400	261		
M>		3 3 C	C 3 3	CAC	mcc		TECC	መመረ	CCG	CAG		CTC	AAC	CCC	GGC		AAC	576	
									Pro									3,0	
W>		пуз	GIU	GIU	405	GIU	Set	FIIC	110	410	пта	пси	цуз	110	415	C _I S			
W>		mcc.	ccc	ጥጥር		TCC	ጥጥር	$C\Delta\Delta$	AAC		GAC	AAC	CCG	TCG		ACC	TTC	624	
									Asn									-	
W>		ııp	пта	420	пор	115	1110	0111	425				110	430					
W >		CAG	GAG		GCC	ጥርር	CCG	ፐርር	GAG	СТС	ACG	ፐርር	AAG		GGC	TGC	·TCC	672	
									Glu										
W>		J111	435	,41		J ₁ J		440					445		1	- 1 -			
,, ,		ССТ		ፐርር	AGC	AGC	ACC		TCT	CCG	GTC	AAC		ССТ	ACC	AGC	ACC	720	
									Ser										
		3																	

PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002
TIME: 11:36:15

W>	269	450					455					460					465	7.50
W>	271	300	ACC	ACG	TCC	ACC	TCC	ACC	ACC	TCG	AGC	CCG	CCA	GTC	CAG	CCT	ACG	768
	272	Ser	Thr	Thr	Ser	Thr	Ser	Thr	Thr	Ser	Ser	Pro	Pro	Val	Gln	PIU	Thr	
W>	272					470					4/5					400		016
., ,	075	ACT	CCC	AGC	GGC	TGC	ACT	GCT	GAG	AGG	TGG	GCT	CAG	TGC	GGC	GGC	AAT	816
	276	Thr	Pro	Ser	Gly	Cys	Thr	Ala	Glu	Arg	\mathtt{Trp}	Ala	Gln	Cys	GLY	GLY	Asn	
W>	~==				405					490					3 / /			864
••	070	GGC	TGG	AGC	GGC	TGC	ACC	ACC	TGC	GTC	GCT	GGC	AGC	ACT	TGC	ACG	AAG	004
	280	Gly	Trp	Ser	Gly	Cys	Thr	Thr	Cys	Vaļ	Ala	Gly	Ser	TIII	Cys	Thr	гàг	
W>	281			500					505					510				894
••	283	ATT	AAT	GAC	TGG	TAC	CAT	CAG	TGC	CTG	TAG							094
	284	Ile	Asn	Asp	Trp	Tyr	His	Gln	Cys	Leu								
W>	285		515					520					•					
	288	(2)	INF	ORMA!	rion	FOR	SEQ	ID i	10: 4	1:								
	290	, ,	(i	SE	QUENC	CE CI	HARAC	TER.	ISTI	CS:								
	291						H: 29			aci	ds						•	
	292			(1	B) T	YPE:	amir	no a	cid				•					
	293			(1	D) T	OPOL	OGY:	line	ear									
	295		(ii) MO	LECU	LE T	YPE:	pro	tein								•	
	297		(xi) SE	QUEN	CE D	ESCR:	[PTI	ON:	SEQ	ID N	0: 4	:	Dwo	w- 1	Tou	λla	•
	299	Met	His	Leu	Ser	Ala	Thr	Thr	Gly	Phe	Leu	Ala	ьeu	Pro	vaı	15	AIG	
	200	1				5					10					10		
	302	Leu	Asp	Gln	Leu	Ser	Gly	Ile	Gly	Gin	Thr	Thr	Arg	Tyr	30	ьэр	Cys	
	303				20				_	25	T	c1	Dro	cor			Val	
	305	Cys	Lys			Cys	Ala	Trp	Pro	GLY	ьys	СТА	PIO	45	Der	110	Val	
	306			35			_		40	Dmo	Tou	λcn	λen		Glv	Ser	Thr	
					Asp	Lys	Asn	Asp	ASI	PIO	Leu	ASII	60	GLI	O ± j	502	Thr	
	309		50		_	•	.1.	55	C1.	Cor	· 7.1 a	ጥህን			Ser	Ser	Gln	
				GLY	Cys	Asp	Ala	СТА	СТУ	361	пти	75		010			80	
	312	65		_		17- 1	70	7 an	Cl u	Τ.Δ.1	Ser			Trp	Ala	Ala	Val	
			Pro	Trp	Ala	va1 85	Ser	АЗР	GIU	БСС	90	-1-	1			95	;	
	315	_		310	C1	. co.	cor	Glu	Ser	Gln			Cvs	Ala	Cys	Туг	Glu	
			ьeu	Ala	100	261	361	GIU		105	_F	-1-	•		110)		
	318		mb.~	Dha	TUU.	Cor	- G1v	Dro	Val	Ala	Glv	Lvs	Lys	Met	Ile	val	Gln	•
	201			115					- 120					123				
	321	1 .	mby	TTO	· · Thr	· 61v	G1v	Asp	Leu	Glv	Asr	Asn	His	Phe	Asp	Let	ı Ala	
	204		120	١				135	1				140	,				
	324	т1.	130	Clu	r Glv	Glv	val	Glv	Ile	Phe	Asi	Ala	Cys	Thr	Asp	Gli	160	
	227	1 4 5	:				150					100)				100	
	220	145	, , אום	Dro	Pro	Asr	Glv	Trr	Gly	Asp	Arg	ту1	Gly	gly	Ile	e His	s Ser	
	220					165	`				1/(,				 /.	,	
	330	T.376	2 (3)	Gli	CVS	Glu	. Ser	Phe	Pro	Glu	ı Ala	Lei	ı Lys	s Pro	Gly	Cy:	s Asn	
	222	,			180	1				18:)				TO	,		
	222	י י∼רידף	a Arc	phe	Agr	Trr	Phe	Glr	.Ası	a Ala	a Asp	Ası	n Pro	ser	· Val	L Th	r Phe	
	226			105	:				200)				203	,			
	336	, R Gli	n Gli	ı Val	L Ala	CVS	Pro	Sei	Glu	ı Leı	ı Thi	r Sei	r Lys	s Ser	Gly	у Су	s Ser	
	220	`	216	`				215)				22	,				
	341	Are	g Pro	sei	c Sei	s Sei	c Thr	: Sei	s Se	r Pro	o Vai	L Ası	a Gl	n Pro	Th	r Se	r Thr	
	741		, . . `															

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/007,521

DATE: 02/04/2002 TIME: 11:36:16

```
L:33 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:34 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:213 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:217 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:221 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:225 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:229 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:233 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:237 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:241 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:245 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:249 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:253 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:257 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:261 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:265 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:269 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:273 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:277 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:281 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:285 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:375 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:379 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:383 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:387 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:391 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:395 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:399 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:403 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:407 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:411 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:415 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:419 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:423 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:427 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:431 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:435 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:439 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:443 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:447 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:451 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:5
L:544 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:548 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:552 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:556 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:560 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:564 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
L:568 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7
```

VERIFICATION SUMMARY

DATE: 02/04/2002

PATENT APPLICATION: US/10/007,521

TIME: 11:36:16

Input Set : N:\Crf3\RULE60\10007521.raw Output Set: N:\CRF3\02042002\J007521.raw

L:572 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7 L:576 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7 L:580 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7 L:584 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:7 L:3513 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:79 L:3529 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:80 L:3545 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:81 L:3561 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:82 L:3577 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:83 L:3907 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:105 L:3922 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:106 L:3937 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:107 L:3952 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:108 L:3967 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:109